Backtracking on the folding landscape of the β-trefoil protein interleukin-1β?

Dominique T. Capraro*, Melinda Roy*, José N. Onuchic†‡, and Patricia A. Jennings*‡

*Department of Chemistry and Biochemistry and †Center for Theoretical Biological Physics, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093

Contributed by José N. Onuchic, August 11, 2008 (sent for review July 24, 2008)

Interleukin-1β (IL-1β) is a cytokine within the β-trefoil family. Our data indicate that the folding/unfolding routes are geometrically frustrated. Follow-up theoretical studies predicted backtracking events that could contribute to the broad transition barrier and the experimentally observed long-lived intermediate. The backtracking route is attributed to the topological frustration introduced by the packing of the functional loop (the β-bulge, residues 47–53) to the nascent barrel. We used real-time refolding NMR experiments to test for the presence of backtracking events predicted from our theoretical studies. Structural variants of IL-1β, a β-bulge deletion, and a circular permutation that opens the protein in the middle of the experimentally observed kinetic intermediate, were also refolded and studied to determine the affects on the observed folding reactions. The functional loop deletion variant demonstrated less backtracking than in WT protein whereas the permutation still maintains backtracking in agreement with theoretical predictions. Taken together, these findings indicate that the backtracking results from geometric frustration introduced into the fold for functional purposes.

protein folding | real-time NMR

The proteins of β-trefoil superfamily have similar native conformations, fold slowly, and have been the subject of numerous experimental and theoretical folding studies (1–10). Interleukin-1β (IL-1β), a cytokine within the β-trefoil family, has a global fold composed of a repeating trefoil unit (βββloop β-motif), forming a six-stranded β-barrel pinned by three two-stranded hairpin turns to form a cap, closing one end of the barrel (3) (Fig. 1 *A* and *B*). The functional loop, the β-bulge, is located between strands 4 and 5, as depicted in Fig. 1. Experimental and theoretical data show that IL-1β folds through an on-route intermediate ensemble characterized by the protection of β-strands 6–10 from hydrogen-deuterium exchange (11–13) (Fig. 1*A*). Real-time NMR kinetic experiments of the denaturant-induced unfolding of IL-1β revealed a rugged landscape where native-like turns were unexpectedly persistent and outlasted all other observed structural features (14).

Interestingly, theoretical refolding studies also indicate the presence of a rugged, broad barrier for the folding kinetics of β-trefoil proteins (1). Three major routes are populated during the refolding of the β-trefoil family, two of which are observed mostly with other members, e.g., the ends-together (FGF-1) and the direct (hisactophilin) routes (Fig. 1*C*). Although IL-1β has not been observed experimentally to fold through the N and C termini ends-together route, we explored whether we could facilitate folding by the ends-together route by linking the N and C termini of WT IL-1β and introducing new terminal ends, thus forming a circular permutation.

Simulations suggest that IL-1β folds by a unique backtracking route not seen in other members of the family. Backtracking is a variation in one of the populated folding routes of IL-1β in which two subsets of contacts compete with each other, resulting in the unfolding of one subset of native contacts and their subsequent refolding later along the folding coordinate (1). It has been suggested that the backtracking route occurs as a result of the topological frustration the functional β-bulge (residues 47–53) causes during folding (2). To test this hypothesis, we removed the β-bulge and compared the results of folding studies for this IL-1β variant to those observed for WT and circular permutant IL-1β proteins.

Although not routine, real-time NMR spectroscopy has been used to follow the unfolding (14–18) and the refolding (19, 20) of a few proteins that fold on an appropriate timescale. Given proper conditions, rapid mixing methods to initiate folding coupled with procedures to reduce the rate of protein folding have made it possible to follow protein folding by NMR on a timescale of seconds to hours, under conditions closely similar to those used for optical techniques (21).

We used real-time NMR methods to track the refolding of WT and mutant IL-1β proteins. In agreement with theoretical predictions (1, 2), we show that IL-1β, a topologically frustrated protein, undergoes backtracking during refolding. In contrast, the mutant protein (the bulge-less), designed to remove the predicted major functional loop responsible for the topological frustration (2), shows little evidence for backtracking behavior in our studies. Taken together, our experimental results support the hypothesis that topological frustration in folding may be introduced by necessary functional moieties and comes at the cost of highly efficient folding.

Results and Discussion

NMR spectroscopy is a useful tool for uncovering the presence of heterogeneous populations representing different folding ensembles of species during folding/unfolding of IL-1β (11, 12, 14). Although monitoring folding events using pulse-labeling hydrogen-deuterium exchange and NMR is an effective method for determining folding events at atomic resolution, the number of proteins that can be probed is limited by the necessity of having sequence-specific resonance assignments and whether the protein can tolerate the necessary sudden changes in pH for pulse-labeling. Oftentimes, aggregation is a problem, even for proteins that are well behaved at neutral pH (22). To circumvent this problem, we chose to monitor refolding in chaotropic conditions where minimal aggregation occurs and the native fold remains intact. Fig. 2 is a heteronuclear single quantum coherence (HSQC) spectrum of WT IL-1β at the completion of the refolding experiment in our final conditions for WT and mutant proteins. The overall chemical shift dispersion in the final refolded conditions of 0.8 M guanidine hydrochloride (Gdn-HCl) is consistent with a well folded protein ensemble (23). Although fast 2D HSQC were sufficiently "quick" to follow real-time unfolding, more time resolution was needed for tracking the folding reaction, in particular, for detecting the presence of backtracking. Hence, we used 1D versions to gather more time

Author contributions: D.T.C., M.R., J.N.O., and P.A.J. designed research, performed research, contributed new reagents/analytic tools, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

‡To whom correspondence may be addressed. E-mail: jonuchic@ucsd.edu or pajennin@ucsd.edu.

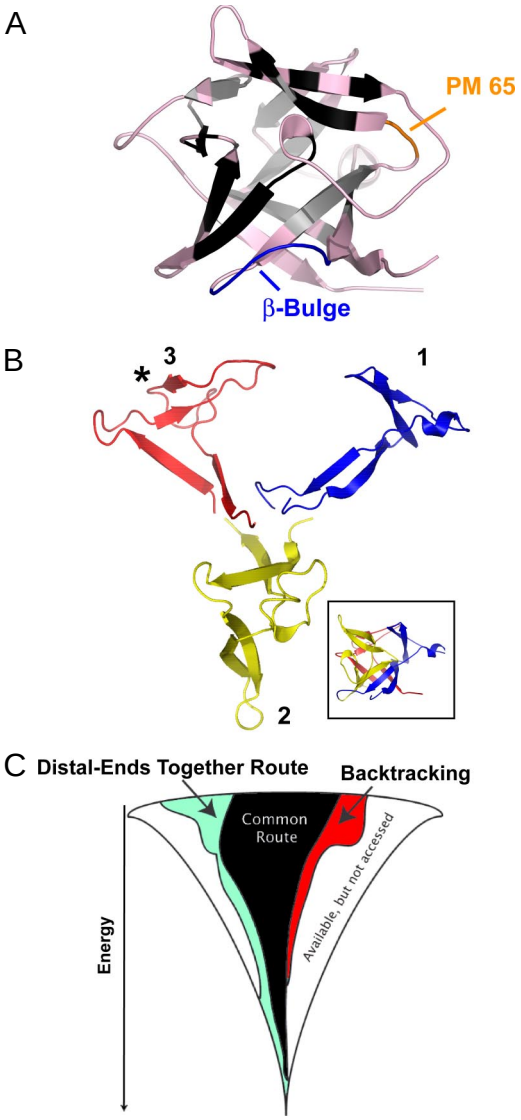

Fig. 1. Representations of the IL-1β and the theoretical folding routes of IL-1β based on simulation data in ref. 1. (*A*) A ribbon diagram of IL-1β demonstrating the areas protected from hydrogen-deuterium exchange and associated with the on-pathway folding intermediate (11). Black indicates areas of early protection, and gray represents areas of late protection. Blue highlights the region in the protein referred to as the β-bulge, and orange demonstrates the location of the cut-site for the circular permutation, PM65. (*B*) The backtracking route where subunit 3 (red) forms first, followed by local unfolding in that region and formation of trefoil 2. Folding then progresses where trefoil subunits 2 (yellow) and 3 (blue) come in, forming the barrel followed by trefoil 1, which leads to the packing of the barrel and the cap into the native structure. For clarity, an asterisk (*) indicates the red trefoil subunit. (*Inset*) The native fold of IL-1β illustrating the location of the trefoil subunits with the same color scheme. (*C*) The energy landscape of the β-trefoil family. The proteins access one main route, initiating in the middle and one flanking trefoil. Nuances of topology or energetics may cause the early packing of the core to be problematic, which would result in the main route being less favorable and two other routes, previously unfavorable, to be accessed. One of these routes brings the distal ends of the protein together first and one backtracks.

points on a given time scale. Using ^{15}N-enriched proteins, we reduced spectral complexity by selecting for amide protons whose chemical shifts reflect the local secondary structure and tertiary contacts of a given amide proton.

Unfolded protein was rapidly introduced in conditions that strongly favor the formation of the native protein and reduce the

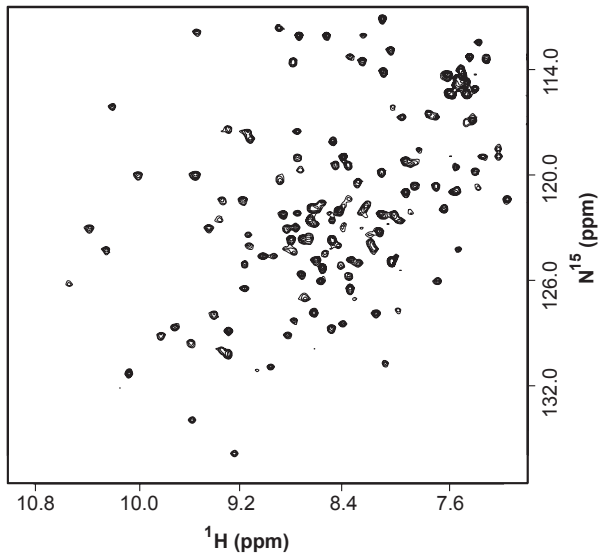

Fig. 2. Representative ^{1}H-^{15}N HSQC spectrum of the final fold of WT IL-1β in 0.8 M Gdn-HCl. A ^{1}H-^{15}N HSQC spectrum of refolded IL-1β in 0.8 M Gdn-HCl at the completion of the refolding reaction is shown. IL-1β was unfolded in 3 M Gdn-HCl, pH 5.4 overnight at 4°C. The refolding reaction was initiated by dilution to a final protein concentration of ≈11 mg/ml and 0.8 M Gdn-HCl. The chemical shift and dispersion of fingerprint region is indicative of intact tertiary structure.

rate of protein folding to a timescale of tens of minutes, to follow refolding by NMR spectroscopy. Once refolding was initiated, backbone amide proton resonances were monitored by the 1D heteronuclear equivalent of a refocused Insensitive Nuclei Enhanced by Polarization Transfer (INEPT) ^{1}H-^{15}N experiment (see *Materials and Methods*). The spectral data resulting from serial acquisition of 1D INEPT spectra during refolding are presented for all IL-1β variants in Figs. 3 and 4. The specific backbone amide proton resonances for residues Asn-119, Asn-102, and His-30 of WT IL-1β are highlighted (labeled or indicated) in Fig. 3*A*. These signals are well resolved and easy to follow over time. The signals appear early during the refolding reaction, followed by fluctuations in intensities over time, until the peak intensities stabilize to those observed for the native protein under the final conditions. This behavior is consistent with the predicted backtracking route from theoretical studies (Fig. 1*B*) in which a subset of the ensemble exhibits a rearrangement within the protein of some of the native contacts along the folding trajectory, causing specific contacts to unfold and refold before the refolding process can be completed (2). This rearrangement has a noticeable effect on the signal intensity of the peaks and is thought to occur in the vicinity of the functionally important β-bulge of IL-1β that is not present in the other family members, even in the highly conserved IL-1β receptor antagonist protein, IL-1ra.

A similar region of the spectrum is given for the β-bulge-less variant in Fig. 3*B* and for the circular permutation in Fig. 3*C*. The series of 1D ^{15}N-HSQC spectra with resonances for Asn-119, Asn-102, and His-30 for the bulge-less variant (Fig. 3*B*) gradually appear and peak intensities increase and then plateau over time. There are significantly reduced fluctuations in intensities compared to those observed for the WT protein (Fig. 3*A*). In contrast, the fluctuations persist in the circular permutant protein (Fig. 3*C*). We attribute these noticeable differences to the backtracking route (1, 2) in which unfolding and refolding of local contacts alter the intensities of the peaks, generating visible differences in the spectra. It is important to note that the appearance and disappearance of the native-like contacts seen in

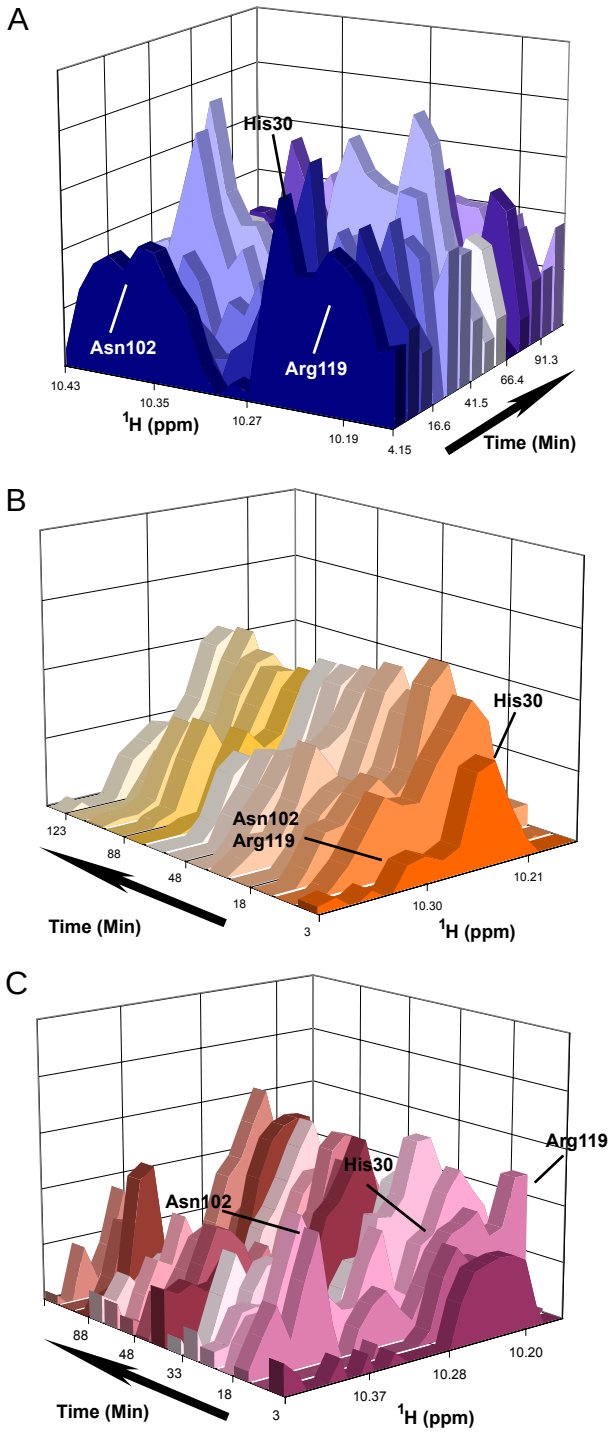

Fig. 3. Time-dependent ^{1}H-^{15}N spectral changes of similar representative regions in WT (blue), the bulge-less construct (orange), and circular permutant (dark red) during refolding depicting evidence of backtracking. (*A*) 1D spectra of resonance peaks for WT IL-1β for residues Arg-119 at 10.18 ppm, His-30 at 10.23 ppm, and Asn-102 at 10.35 ppm. The black arrow indicates increasing time from the initiation of refolding to the completion of the refolding. (*B*) 1D spectra representing the same three amino resonances for the bulge deletion construct for residues His-30 at 10.24 ppm and the overlapping peak of Asn-102 and Asn-119 at 10.30 ppm. The black arrow indicates increasing time from the initiation of refolding to the completion of the refolding. (*C*) 1D spectra representing the same three amino resonances for the circular permutant for residues Arg-119 at 10.18 ppm, His-30 at 10.23 ppm, and Asn-102 at 10.35 ppm. The black arrow indicates increasing time from initiation of refolding to completion of refolding. For clarity, the orientation of the stacked spectra is chosen for maximal clarity.

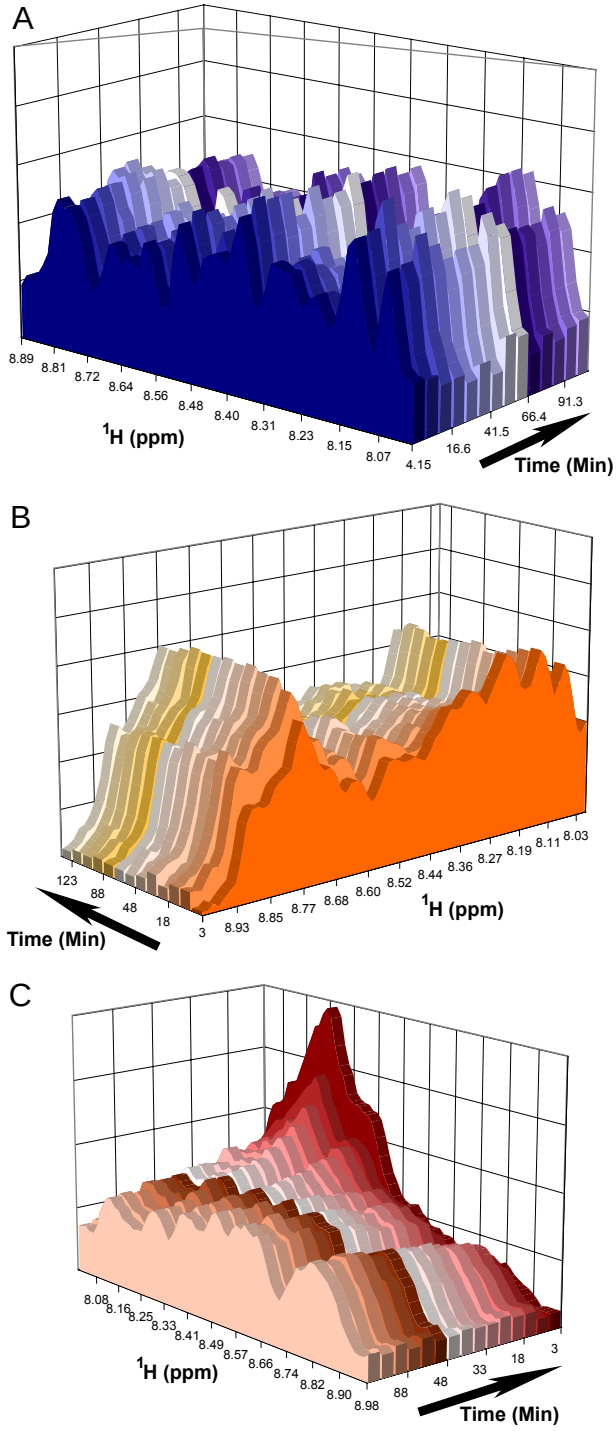

Fig. 4. Time-dependent ^{1}H-^{15}N spectral changes of similar representative regions in WT (blue), the bulge-less construct (orange), and circular permutant (dark red) during refolding depicting global refolding. The 1D spectra of resonance peaks for WT IL-1β (*A*), bulge-less variant (*B*), and permutant (*C*) proteins between 8 and 9 ppm are shown. The black arrow indicates increasing time from the initiation of refolding to the completion of the refolding. The series of spectra demonstrates a simple progression of refolding as peak intensities strengthen, illustrating the global refolding process. The orientation of the stacked spectra is chosen for maximal clarity.

WT (Fig. 3*A*) does not reflect the transient population and depopulation of a common species. Rather, the contacts, although native-like, represent partial disruption within a larger ensemble of contacts in various species along the folding path-

way. These disruptions allow the observation of an oscillating appearance of contacts in a folding pathway that has an otherwise smooth trajectory. That is, the backtracking observed reflects a local, not global, phenomenon. Representations of the same regions of the spectra as a function of time are characterized by the steady formation of resonance peaks increasing over time. The steady progression to natively folded protein is apparent for the WT and mutant proteins (Fig. 4) and correlates with theoretical results, which suggest that only particular subregions of IL-1β exhibit backtracking. Taken together with theoretical predictions (1, 2), these results are consistent with the relationship between the geometrical frustration introduced by the β-bulge and the backtracking route.

Although family members may prefer a particular folding route, small changes in solution conditions alter the associated energies (i.e., caused by changes in pH, mutations, ionic strength, pressure, etc.) and thus may result in proteins folding along different paths. This feature of the energy landscape leads to a certain robustness that still facilitates folding under diverse situations. However, introduction of specific functional units that lead to geometric frustration during folding can alter the energy landscape and impede fast folding. Fig. 1*C* is a cartoon representation of the energy landscape observed for the folding of the β-trefoil family of proteins from simulated data (1, 2). Interestingly, hisactophilin, the simplest of the β-trefoil family proteins studied to date, folds the most efficiently and by the direct route, where the trefoil subunits fold in progression of one another (3–2-1 or 1–2-3). However, introduction of specific contacts in the simulations can lead to frustration in folding, not unlike that seen for the IL-1β protein. Because topology determines which order of contact formations leads to the native state most efficiently, the chain connectivity and the distribution of contact lengths can make part of the energy landscape inaccessible and lead to a rerouting beyond a certain point (2, 24–29). As a result of linking the terminal ends of WT IL-1β and introducing a new N and C termini, forming a circular permutation, the refolding NMR data indicate a noticeably slower progression to a fully refolded protein, while still sampling the backtracking route. The presence of the route is consistent with simulations showing that perturbations in the energetics of IL-1β altered the dominant folding route. The overall landscape remained the same but the dominant route shifted to those routes previously less sampled (1) and suggests that protein folding would persist if the backtracking route could be blocked. Examples of this behavior have also been seen with bovine pancreatic trypsin inhibitor (BPTI). Local structural rearrangements have been observed during the refolding of BPTI that contains a disulfide bond in addition to a bulge loop (30). Alteration in the chain connectivity around the disulfide resulted in altered folding routes, while the folding landscape remained similar (31).

Conclusion

Theoretical data show that IL-1β refolds from an unfolded state, sampling a unique folding route that backtracks and partially unfolds the intermediate to accommodate the functional loop in addition to sampling similar folding routes attributed to the β-trefoil family. The backtracking route is attributed to the accommodation of the functional loop, introducing a functional aspect to the folding landscape. Disruption of the energetics of WT IL-1β "reroutes" the folding, changing the sampling of the various folding routes accessed by the protein. Removal of the functional loop and rearrangement of the primary sequence both alter the observed route of refolding. These experimental findings indicate backtracking being observed during folding and are in agreement with the theoretical models. This suggests that the landscape of the funnel is robust, and that the protein is able to fold even if the route most easily traversed is blocked by small changes to the protein or its environment (1).

Materials and Methods

Cloning, Expression, and Protein Purification. Cloning, expression, purification, and purity of ^{15}N-enriched WT IL-1β and IL-1β variants followed published protocols (14). Cloning, expression, and purification of the deletion construct of IL-β, which has the β-bulge removed (residues 47–53) and replaced with a three-residue turn, Lys-47, His-48, and Ser-49, was purified in pH 6.5 buffers, following WT purification protocol. The circular permutation of IL-1β also followed similar cloning, expression, and purification protocols as WT IL-1β.

Real-Time Refolding via NMR. ^{15}N-enriched protein samples were unfolded in 100 mM sodium actetate-d_3 (Aldrich), 3 M guanidine hydrochloride (Gdn-HCl), pH 5.4 overnight. Refolding was initiated by dilution of the unfolded protein solution with 100 mM sodium acetate-d_3, 5% D_2O to a final concentration of 0.8 M Gdn-HCl for WT IL-β and the bulge-less construct, and 0.6 M Gdn-HCl for the circular permutation, and a final protein concentration of $\approx$4 mM. Experiments were performed at 22°C.

All NMR spectroscopy on ^{15}N-enriched IL-1β, the bulge-less and circular permutant proteins were obtained on a Bruker DMX 500 NMR spectrometer, equipped with a triple-resonance gradient probe. 2D ^{1}H-^{15}N HSQC spectra were acquired as before (14, 32).

To acquire more data points during refolding, 1D ^{1}H spectra were serially acquired instead of 2D HSQC as in ref. 14. To simplify the crowded protein spectrum, we used the 1D heteronuclear equivalent of a refocused INEPT ^{1}H-^{15}N experiment, which results in a spectrally edited spectrum (33). INEPT transfers the magnetization from the ^{1}H nuclei to the ^{15}N and back again to ^{1}H, resulting in a proton spectrum of only the amide protons resonances and those of the side-chain NH_2. NH_2 resonance signals originating from residual denaturant, Gdn-HCl, were reduced by a selective 180° shifted Laminar pulse. The offset was set at $-$978 Hz and a Sinc1.1000 shape was used for the pulse. Pulse amplitude or power was selected to achieve the lowest intensity for the guanidine peak and maximize amide proton signal intensities.

Sixty 1D INEPT spectra were collected after initiation of protein refolding with a dead time of 2 min. Individual spectrum acquisition time was 3 min, and total run time was 182 min. Sequence-specific assignments for amide proton resonances were based on published work (14) and 3D CBCA(CO)NH and HNCA spectra (34). 2D ^{1}H-^{15}N HSQC spectra were processed with Felix 2004 software (MSI), and 1D INEPT spectra were processed with MetReC-Lite (Mestrelab Research).

ACKNOWLEDGMENTS. We thank Xuemei Huang and Anthony Mrse for assistance with the NMR experiments, the Keck Institute (Claremont, CA) for computational support, and members of the laboratories of J.N.O. and P.A.J. for discussion. This work was funded by National Institutes of Health Grant GM54038 (to P.A.J. and J.N.O.).

1. Chavez LL, Gosavi S, Jennings PA, Onuchic JN (2006) Multiple routes lead to the native state in the energy landscape of the β-trefoil family. *Proc Natl Acad Sci USA* 103:10254–10258.
2. Gosavi S, Chavez LL, Jennings PA, Onuchic JN (2006) Topological frustration and the folding of interleukin-1β. *J Mol Biol* 357:986–996.
3. Murzin AG, Lesk AM, Chothia C (1992) β-Trefoil fold. Patterns of structure and sequence in the Kunitz inhibitors interleukins-1β and 1α and fibroblast growth factors. *J Mol Biol* 223:531–543.
4. Murzin AG, Lesk AM, Chothia C (1994) Principles determining the structure of β-sheet barrels in proteins. II. The observed structures. *J Mol Biol* 236:1382–1400.
5. Dubey VK, Lee J, Blaber M (2005) Redesigning symmetry-related "mini-core" regions of FGF-1 to increase primary structure symmetry: Thermodynamic and functional consequences of structural symmetry. *Protein Sci* 14:2315–2323.
6. Murray KB, Taylor WR, Thornton JM (2004) Toward the detection and validation of repeats in protein structure. *Proteins* 57:365–380.
7. Liu C, Gaspar JA, Wong HJ, Meiering EM (2002) Conserved and nonconserved features of the folding pathway of hisactophilin, a β-trefoil protein. *Protein Sci* 11:669–679.
8. Liu C, *et al.* (2001) Thermodynamics of denaturation of hisactophilin, a β-trefoil protein. *Biochemistry* 40:3817–3827.
9. Srisailam S, *et al.* (2003) Amyloid-like fibril formation in an all β-barrel protein: Partially structured intermediate state (s) is a precursor for fibril formation. *J Biol Chem* 278:17701–17709.
10. Chi Y, *et al.* (2001) Thermodynamic characterization of the human acidic fibroblast growth factor: Evidence for cold denaturation. *Biochemistry* 40:7746–7753.

11. Heidary DK, Gross LA, Roy M, Jennings PA (1997) Evidence for an obligatory intermediate in the folding of interleukin-1β. *Nat Struct Biol* 4:725–731.
12. Varley P, *et al.* (1993) Kinetics of folding of the all β-sheet protein interleukin-1-β. *Science* 260:1110–1113.
13. Clementi C, Jennings PA, Onuchic JN (2000) How native-state topology affects the folding of dihydrofolate reductase and interleukin-1β. *Proc Natl Acad Sci USA* 97:5871–5876.
14. Roy M, Jennings PA (2003) Real-time NMR kinetic studies provide global and residue-specific information on the noncooperative unfolding of the β-trefoil protein, interleukin-1β. *J Mol Biol* 328:693–703.
15. Kiefhaber T, Labhardt AM, Baldwin RL (1995) Direct Nmr evidence for an intermediate preceding the rate-limiting step in the unfolding of ribonuclease A. *Nature* 375:513–515.
16. Akasaka K, Naito A, Imanari M (1991) Novel method for Nmr spectral correlation between the native and the denatured states of a protein: Application to ribonuclease A. *J Am Chem Soc* 113:4688–4689.
17. Adler M, Scheraga HA (1988) Structural studies of a folding intermediate of bovine pancreatic ribonuclease A by continuous recycled flow. *Biochemistry* 27:2471–2480.
18. Frieden C, Hoeltzli SD, Ropson IJ (1993) Nmr and protein folding: Equilibrium and stopped-flow studies. *Protein Sci* 2:2007–2014.
19. Koide S, Dyson HJ, Wright PE (1993) Characterization of a folding intermediate of apoplastocyanin trapped by proline isomerization. *Biochemistry* 32:12299–12310.
20. Schanda P, Forge V, Brutscher B (2007) Protein folding and unfolding studied at atomic resolution by fast two-dimensional NMR spectroscopy. *Proc Natl Acad Sci USA* 104:11257–11262.
21. Balbach J, *et al.* (1995) Following protein folding in real time using Nmr spectroscopy. *Nat Struct Biol* 2:865–870.
22. Thirumalai D, Klimov DK, Dima RI (2003) Emerging ideas on the molecular basis of protein and peptide aggregation. *Curr Opin Struct Biol* 13:146–159.
23. Roy M, *et al.* (2005) The native energy landscape for interleukin-1β. Modulation of the population ensemble through native-state topology. *J Mol Biol* 348:335–347.
24. Plaxco KW, Simons KT, Baker D (1998) Contact order, transition state placement, and the refolding rates of single domain proteins. *J Mol Biol* 277:985–994.
25. Chiti F, *et al.* (1999) Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. *Nat Struct Biol* 6:1005–1009.
26. Otzen DE, Fersht AR (1998) Folding of circular and permuted chymotrypsin inhibitor 2: Retention of the folding nucleus. *Biochemistry* 37:8139–8146.
27. Koga N, Takada S (2001) Roles of native topology and chain-length scaling in protein folding: A simulation study with a Go-like model. *J Mol Biol* 313:171–180.
28. Fersht AR (2000) Transition-state structure as a unifying basis in protein-folding mechanisms: Contact order, chain topology, stability, and the extended nucleus mechanism. *Proc Natl Acad Sci USA* 97:1525–1529.
29. Thirumalai D, Klimov DK (2007) Intermediates and transition states in protein folding. *Methods Mol Biol* 350:277–303.
30. Goldenberg DP, Creighton TE (1983) Circular and circularly permuted forms of bovine pancreatic trypsin inhibitor. *J Mol Biol* 165:407–413.
31. Bulaj G, Koehn RE, Goldenberg DP (2004) Alteration of the disulfide-coupled folding pathway of BPTI by circular permutation. *Protein Sci* 13:1182–1196.
32. Heidary DK, Roy M, Daumy GO, Cong Y, Jennings PA (2005) Long-range coupling between separate docking sites in interleukin-1β. *J Mol Biol* 353:1187–1198.
33. Miao XJ, Freeman R (1995) A spin-echo technique for separation of multiplets in crowded spectra. *J Magn Reson Ser A* 116:273–276.
34. Cavanagh J (1996) *Protein NMR Spectroscopy: Principles and Practice* (Academic, San Diego).